Chapter 4

Noise and Sensitivity in Interferometry

Charles H. Townes

University of California at Berkeley
Berkeley, California

4.1 Introduction

The quality of an interferometer may be measured by a number of parameters—these include the precision of visibility measurements, sensitivity in measuring weak sources, precision in measuring fringe phases, range of baselines available, and general flexibility. Some of these factors will be discussed here, with emphasis on noise phenomena and their effect on sensitivity.

4.2 Wavefront Aberrations

Although complete interference can occur if the wavefronts are both perfectly planar as they strike the two telescopes, they are likely to be distorted over a given telescope aperture by imperfect seeing, which can give misleadingly low values of visibility. Heterodyne detection selects and detects only the components of the wavefront of the stellar radiation that are in phase with the wavefront of the laser local oscillator (cf. Kingston 1978), and thus tends to prevent this difficulty. A similar result can be achieved when direct, rather than heterodyne, detection is used by spatial filtering with a glass fiber to obtain a single geometric mode. However, this usually entails some loss of signal.

4.3 Thermal and Quantum Noise

4.3.1 Noise Power Fluctuations

The fundamental noise for an ideal direct detector, which detects both polarizations, is due to thermal radiation striking the detector, which for an ideal detector produces a noise power fluctuation of

$$N_d = h\nu\sqrt{\frac{2\Delta\nu}{t}\frac{1-\varepsilon}{e^{h\nu/kT}-1}}, \tag{4.1}$$

where $\Delta\nu$ is bandwidth, t the averaging time, T the temperature of optics and atmosphere through which the signal is received, and ε the fractional transmission of radiation reaching the telescope. This expression is valid for a photodiode; photoconductors have more noise by a factor of $\sqrt{2}$. Direct detection does not determine the phase of a wave, and hence noise due to the uncertainty principle is not present; in principle, noise is due only to fluctuations in the number of quanta in the radiation received. These fluctuations are also present in heterodyne detection, but at IR frequencies are generally much smaller than the uncertainty principle noise, and hence are omitted from Equation 4.2 below, where heterodyne detection is discussed.

The fundamental noise power for a heterodyne detector, (which detects only one polarization, i.e., that of the local oscillator) is equivalent to an average of one quantum per second per unit bandwidth in the same polarization as the local oscillator. For an ideal photodiode, the noise power fluctuation is hence

$$N_h = h\nu\sqrt{\frac{2\Delta\nu}{t}}, \tag{4.2}$$

where $h\nu$ is the quantum energy, $\Delta\nu$ is the single sideband bandwidth in Hz, and t the post-detection averaging time in seconds (cf. Teich 1970, Kingston 1978, Townes 1984). Since heterodyne detection has the ability to measure the phase of a wave, and phase is complementary to energy or number of quanta, this noise is an inescapable result of quantum mechanics and the uncertainty principle (cf. Serber and Townes 1960; Kimble and Walls 1987). The ratio of uncertainties in number of photons to those in phase of the wave can be changed while still satisfying the uncertainty principle (ibid.), but such possibilities are not very practical for heterodyne detectors and are not considered here.

4.3.2 Signal-to-Noise Ratio

For a useful signal, equivalent power of noise fluctuations must be substantially less than the signal power P_ν from a source observed. The signal-to-noise ratio for the two cases, assuming ideal detectors with 100% quantum efficiency, is

$$(S/N)_h = \frac{P_\nu}{h\nu}\sqrt{2\Delta\nu t}, \tag{4.3}$$

and

$$(S/N)_d = \frac{P_\nu}{h\nu}\sqrt{\frac{2\Delta\nu t\left(e^{h\nu/kT}-1\right)}{1-\varepsilon}}, \tag{4.4}$$

where P_ν is the power in each polarization per unit bandwidth (Hertz). Thus, even when direct detection has bandwidths as narrow as heterodyne detection, it appears to have a substantial advantage by a factor:

$$\sqrt{\frac{e^{h\nu/kT}-1}{1-\varepsilon}}, \tag{4.5}$$

and indeed it does under some circumstances. For a wavelength of 10 μm, room temperature T of 293 K, and transmission ε=0.9 (characteristic of atmospheric transmission at 10 μm), this factor can be as large as 37. For visible or near-IR radiation the sensitivity advantage of direct detection is usually overwhelming, even though direct-detection interferometers have complex enough optics that the net transmission is usually substantially less than 0.9—sometimes as low as 0.05. However, there are a number of other considerations, which for wavelengths as large as 10 μm can in some cases give a substantial advantage to heterodyne detection, and in other cases to direct detection. These will be discussed below.

4.4 Heterodyne versus Direct-Detection Interferometry

4.4.1 Signal-to-Noise Ratio for Fringe Measurements

The signal-to-noise for measurement of fringe power in an interferometer with heterodyne detection is (Townes, 1984; cf. also Johnson, 1974, for detailed discussion of detection with a photoconductor)

$$(S/N)_{\text{Fringe}} \equiv \left(\frac{\text{vis}\times P_\nu}{h\nu}\right)^2 (t_0 t)^{1/2}\,\Delta\nu, \tag{4.6}$$

where

P_ν $\equiv$ the power per unit bandwidth of source, as in Equation 4.3
$h\nu$ $\equiv$ the quantum energy
vis $\equiv$ the visibility or fraction of the source power which provides interference
t_0 $\equiv$ the total observing time
t $\equiv$ the length of time atmospheric fluctuations do not change the fringe phase by more than approximately one radian.
$\Delta\nu$ $\equiv$ the single sideband IF bandwidth, as in Equation 4.3

The signal-to-noise for fringe amplitude, and hence for visibility determination, is proportional to the square root of that for fringe power.

The signal-to-noise for a direct-detection interferometer involves somewhat more complex considerations. This is because the relative pathlengths for the signals from the two telescopes must be accurately tracked, and this tracking usually involves radiation of a different wavelength and bandwidth from that used for visibility measurements. If the relative pathlengths can be tracked accurately, then the theoretical signal-to-noise for fringes is comparable to that for power, but reduced by the visibility. It can in principle be made large by long-term averaging. However, the pathlengths themselves fluctuate quite rapidly on timescales of roughly 0.01 to 1.0 seconds due to atmospheric seeing, and there must be enough sensitivity to determine the phase of interference between light from the two telescope sources during a time as short as these fluctuations. If the signal-to-noise approaches unity or less for these short times, no measurement of fringe intensity can be made. Long-term observations for averaging and improving the signal-to-noise are then also not possible.

Equations 4.4 and 4.5 assume essentially perfect detectors and optics. Actually, detectors normally have quantum efficiencies in the range 0.2 to 0.8, which reduces sensitivity by these factors. In addition, transmission of signals through the optical systems involves losses, which for stellar interferometers typically reduces signals by factors between 0.05 and 0.8, depending on the number of mirrors or other optical components involved.

Heterodyne detection sensitivity actually obtained on interference fringes is rather close to the theoretical limit given above. For fringe amplitude, the Infrared Spatial Interferometer (ISI) is within a factor of about 4 of this limit, with much of this factor due to present detector quantum efficiencies being between about 0.25 and 0.40 rather than unity. This closeness to the theoretical limit is in part because heterodyne detection conveniently eliminates unwanted radiation outside a chosen bandwidth and partly because only a very small source at a long distance can produce interference, so more local and extraneous radiation does not provide a false interference signal. However, in contrast to fringe observation, the detection of power radiated into a single telescope suffers from variations in stray radiation as does direct detection, discussed below.

4.4.2 Bandwidth Considerations

For bandwidths narrower that about 1 cm^{-1} (3×10^{10} Hz), the dominant noise in direct detection is typically associated with detector dark current and readout fluctuations (or "read" noise), combined with some stray radiation. For narrow bandwidths, the highest sensitivity normally achieved in direct-detection ground-based astronomical systems is $\sim 2 \times 10^{-15}$ W for a one second averaging time.* This equals the theoretical noise for heterodyne detection with a bandwidth of about 5×10^9 Hz. Of course, heterodyne detection also does not give perfect theoretical performance, largely because the quantum efficiency of detectors is not 100%. However, its noise does decrease with decreasing bandwidth in accordance with theory rather than reaching a lower limit, as is characteristic of direct

*J.H. Lacy, private communication, 1998. Also personal experience of C.H. Townes.

detection. Hence, for bandwidths appreciably narrower than about 5×10^9 Hz, heterodyne detection is typically the more sensitive, since its noise continues to decrease with decreasing bandwidth. This is important for certain applications such as measuring spectral lines (which are often as narrow as 5×10^7 Hz).

For continuum radiation, direct detection has the advantage of being able to use broad bandwidths. The single-sideband bandwidths for efficient heterodyne detectors at 10 μm are presently ~ 0.1 cm^{-1}, or 3×10^9 Hz. However, quantum-well detectors have been made with substantially larger bandwidths, and those as large as ~ 0.5 cm^{-1} can be envisioned. Direct detection can in principle include essentially all the mid-IR radiation transmitted by the atmosphere, or a bandwidth as large as 600 cm^{-1}. However, without further division of the radiation such a broad band of wavelengths results in a rather broad range of resolutions, with varying visibilities. Hence a range of not more that 10%, which is 1 μm in wavelength range or 100 cm^{-1}, is the maximum considered here. A bandwidth of 10–20 cm^{-1} would be normal, giving a 1–2% range of resolution. Such bandwidths provide a substantial apparent advantage over the narrower-band heterodyne detection. However, for broad bandwidths, fluctuations in radiation from the sky and optics usually dominate the noise rather than fundamental quantum fluctuations or detector noise. For a 1-μm bandwidth, experience shows that non-fundamental noise prevents the detection of power below about 5×10^{-14} W for an averaging time of 1 s.† This is about 40 times larger than what would be obtained with the theoretical limit given above, and for this case much of the expected gain from broad bandwidths is negated. In principle, the sensitivity may possibly come closer to theoretical values, but this represents a long-standing challenge to experimentalists.

4.4.3 Arrays with Multiple Telescopes

Interferometers frequently use multiple telescopes to obtain many baselines simultaneously; for example the VLA radio interferometer uses 27 telescopes (or 351 baselines). This requires sending individual signals from each telescope to 26 different interference measurements or correlators. For heterodyne detection, such a system involves no further loss in signal-to-noise, because after detection the signal can be amplified and divided without introduction of any significant noise, as is done in radio interferometry. However, a direct-detection interferometer would normally divide the signal of each telescope into 26 equal parts. This would reduce the signal by the same factor and negate much of the theoretical signal-to-noise advantage noted above. In addition, bringing the signals together from separate telescopes for correlation is also more tractable in the case of heterodyne detection, requiring only electrical cables rather than the evacuated light-pipes and multiple optical

†For 10% bandwidth at 11.7 μm and 1-s averaging time, the MIRAC2 camera used on the UKIRT telescope has a sensitivity (1σ uncertainty) of 5×10^{-13} W (Hoffman et al. 1998). For 1-μm spectral width and 1-s averaging time, specification for the MICS mid-IR camera on the UKIRT telescope is 1.2×10^{-13} W, and for 0.1-μm spectral width, it is 1.3×10^{-14} for a 1-s averaging time (Miyata et al., 1999). When used as a 10-μm camera on the Keck telescope, the LWS has a sensitivity 5.9×10^{-14} W for a 1-μm bandwidth and 1-s averaging time (Keck website LWS Instrument Document).

components which are needed for direct detection. These considerations lead us to conclude that a many-telescope multiple baseline system for the mid-IR region is probably simpler, cheaper, and more flexible with heterodyne detection than with direct detection.

4.4.4 Coherence Time and Fringe Tracking

A prominent task for direct-detection interferometry is to accurately track the relative delay between any two beams which are to interfere. As noted above, this is generally done at a wavelength different from that used for fringe measurement and with a relatively broad bandwidth. However, a useful signal for delay compensation must be obtained in a time as short as atmospheric fluctuations, hence in about 0.01 s. If this signal is detected well, this short time has no direct effect on the sensitivity of measurement, and the delay compensation is helpful in decreasing noise because the phase of interference does not then fluctuate, as it may for heterodyne detection. However, if the signal is lost, visibilities cannot be measured. In contrast, the narrow bandwidth and IF delays of heterodyne detection allow the use of calculated delays without direct tracking, and hence the time for averaging a signal can be arbitrarily long regardless of signal strength. This is another basic convenience of narrow bandwidths. For an averaging time of one hour instead of 0.01 s, this can boost the relative advantage of heterodyne interferometry in detecting fringe power by as much as 600 (cf. Equation 4.6), or in detecting fringe amplitude by a factor of $\sqrt{600} \simeq 25$. The averaging time can also be increased to many hours by many nights of observing, as is often done in radio astronomy. Direct detection can avoid this disadvantage, however, if there is a bright star close enough to be within the same isoplanatic patch as the object being measured, since then the bright star can be used for delay line tracking. And there are some other techniques, not yet generally used, which can ameliorate this sometimes large disadvantage of direct detection. These can involve use of multiple wavelength bands, each possibly as large as atmospheric dispersion allows, to track the pathlength variations, or multiple tracking units, each involving very short distances.

4.5 Conclusion

If a relatively wide bandwidth and single baseline are used, and if there is a strong guide star within the same isoplanatic patch as the object observed, then for wavelengths as short as 10 μm direct detection has a large advantage. However, if many baselines are used and there is no strong guidestar, or if bandwidths less than about 1 cm^{-1} are used, heterodyne detection has a substantial advantage. Both techniques are useful, each has its own optimum functions, and achievement of ideal performance in either one is challenging.

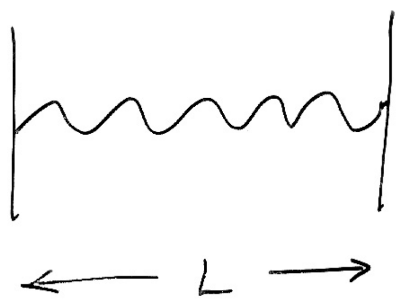
Various expressions and factors discussed above are outlined in the following brief notes and tables.

1.

<u>number of quanta per mode</u> in thermal equilibrium

$$\bar{n} = \frac{\sum_{m=0}^{\infty} m\, e^{-\frac{mh\nu}{kT}}}{\sum_{m=0}^{\infty} e^{-\frac{mh\nu}{kT}}} = \frac{1}{e^{h\nu/kT}-1}$$

<u>Power flow</u>



$$\frac{c}{\nu} = \lambda = \frac{2L}{n} \qquad \therefore \quad n = \frac{2L}{c}\nu$$

Assume 1 photon per mode, Then $\Delta n = \frac{2L}{c}\Delta\nu$

Round trip time $t = \frac{2L}{c}$ $\therefore$ $\frac{\Delta n}{t} = \Delta\nu$

or, for 2 polarizations, the total photon flux $= 2h\nu\Delta\nu$ or, at thermal equilibrium $= \frac{2h\nu\Delta\nu}{e^{\frac{h\nu}{kT}}-1}$

Figure 4.1: The photon flux per unit mode or frequency interval.

2.

For fractional loss ε in path to $T \to 0$,

$$\text{photon flux} = \frac{2 h\nu \Delta\nu \varepsilon}{e^{h\nu/kT} - 1}$$

For signal power $P_\nu \Delta\nu$ in each polarization

$$S/N = \frac{2 P_\nu \Delta\nu t}{h\nu \sqrt{\frac{2 \Delta\nu \varepsilon t}{e^{h\nu/kT} - 1}}}$$

$$= \frac{P_\nu}{h\nu} \sqrt{\left(e^{h\nu/kT} - 1\right) \frac{2 \Delta\nu t}{\varepsilon}}$$

where t is time length of measurement in seconds.

N.B. This is for a photodiode
For a photoconductor, S/N is less by $\sim \frac{1}{\sqrt{2}}$

Figure 4.2: Ideal signal-to-noise ratio for direct detection.